

Causal Inference на Python

Причинно-следственные связи в IT-разработке

Матеуш Факур



Санкт-Петербург • Москва • Минск

2025

Выпущено
при поддержке

КРОК

Краткое содержание

ЧАСТЬ I

Основы

Глава 1. Знакомство с причинно-следственным анализом28

Глава 2. Рандомизированные эксперименты и статистический анализ.....57

Глава 3. Графовые причинные модели.....88

ЧАСТЬ II

Корректировка смещения

Глава 4. Неожиданная эффективность линейной регрессии 124

Глава 5. Мера склонности 172

ЧАСТЬ III

Гетерогенность эффектов и персонализация

Глава 6. Гетерогенность эффектов..... 210

Глава 7. Метаобучаемые системы..... 238

ЧАСТЬ IV

Панельные данные

Глава 8. Разность разностей..... 264

Глава 9. Синтетический контроль..... 305

ЧАСТЬ V

Альтернативные варианты дизайна эксперимента

Глава 10. Гео- и switchback-эксперименты 338

Глава 11. Некомплаентность и инструменты 365

Глава 12. Что дальше 390

Оглавление

Предисловие	16
Для кого эта книга	18
Структура книги	20
Условные обозначения	21
Использование исходного кода примеров	22
Благодарности	23
О научном редакторе русского издания	25
От издательства	26

ЧАСТЬ I

ОСНОВЫ

Глава 1. Знакомство с причинно-следственным анализом	28
Что такое причинно-следственный анализ	28
Для чего нужен причинно-следственный анализ	29
Машинное обучение и причинно-следственный анализ	30
Корреляция и причинно-следственная связь	31
Воздействие и результат	33
Фундаментальная проблема причинно-следственного анализа	33
Причинные модели	35
Вмешательство	37
Эффект индивидуального воздействия	38
Потенциальные результаты	39

Согласованность и предположение о стабильности воздействия	40
Параметры причинности	40
Параметры причинности: пример.....	42
Смещение.....	45
Формула смещения	46
О смещении наглядно	47
Идентификация эффекта воздействия	50
Предположение о независимости.....	52
Идентификация с рандомизацией.....	52
Основные выводы.....	56
Глава 2. Рандомизированные эксперименты и статистический анализ.....	57
Брутфорс-независимость с рандомизацией	57
Пример А/В-тестирования	59
Идеальный эксперимент.....	63
Самая опасная формула	64
Стандартная ошибка оценивания.....	67
Доверительные интервалы.....	69
Проверка гипотез.....	76
Нулевая гипотеза	78
Статистические критерии	80
Р-значения	81
Мощность.....	83
Вычисление объема выборки	85
Основные выводы.....	86
Глава 3. Графовые причинные модели	88
Подход к причинности.....	88
Наглядное представление причинных связей	90
Стоят ли консультанты расходов на них?	93

Краткий обзор графовых моделей	94
Цепи	94
Вилки	96
Перевернутая вилка, или коллайдер	98
Шпаргалка по логике перемещения зависимостей	99
Запросы к графу в Python	100
Снова об идентификации	103
СИА и формула корректировки	104
Предположение о положительности	106
Пример идентификации с данными	106
Ошибка конфаундинга	108
Суррогатный конфаундинг	109
И снова о рандомизации	110
Смещение выборки	111
Введение условия по коллайдеру	112
Корректировка смещения выборки	116
Введение условия по посреднику	119
Основные выводы	120

ЧАСТЬ II

КОРРЕКТИРОВКА СМЕЩЕНИЯ

Глава 4. Неожиданная эффективность линейной регрессии	124
Все, что вам нужно, это линейная регрессия	124
Для чего нужны модели	125
Регрессия при A/B-тестировании	127
Коррекция с применением регрессии	129
Теория регрессии	133
Линейная регрессия с одной переменной	135
Многомерная линейная регрессия	135

Теорема Фриша — Во — Ловелла и ортогонализация	136
Устранение смещения	137
Шумоподавление	139
Стандартная ошибка оценки регрессии	140
Итоговая модель результата	141
Основные выводы по FWL	142
Регрессия как модель результата	144
Положительность и экстраполяция	146
Нелинейные зависимости в линейной регрессии	148
Линеаризация воздействия	149
Нелинейная теорема FWL и устранение смещения	151
Регрессия для «чайников»	152
Условно случайные эксперименты	153
Фиктивные переменные	155
Насыщенная регрессионная модель	158
Регрессия как средневзвешенное значение дисперсии	160
Исключение среднего и фиксированные эффекты	162
Смещение пропущенных переменных: конфаундинг с точки зрения регрессии	165
Нейтральные контрольные переменные	166
Контрольные переменные, создающие шум	168
Отбор признаков: компромисс между смещением и дисперсией	169
Основные выводы	171
Глава 5. Мера склонности	172
Влияние тренингов для менеджеров	172
Корректировка с использованием регрессии	174
Мера склонности	175
Оценка меры склонности	176
Мера склонности и ортогонализация	177

Сопоставление меры склонности.....	178
Обратное взвешивание меры склонности	180
Дисперсия IPW	183
Стабилизированные веса меры склонности	186
Псевдовыборки.....	188
Смещение выборки	189
Компромисс между смещением и дисперсией	190
Положительность	191
Идентификации на основе дизайна и модели	193
Двойное робастное оценивание	194
Простое моделирование воздействия.....	196
Простое моделирование результата.....	199
Обобщенная мера склонности (GPS) для непрерывного воздействия.....	200
Основные выводы	207

ЧАСТЬ III

ГЕТЕРОГЕННОСТЬ ЭФФЕКТОВ И ПЕРСОНАЛИЗАЦИЯ

Глава 6. Гетерогенность эффектов	210
От ATE к CATE	210
Почему прогнозирование не даст ответа	211
CATE с регрессией	214
Оценка прогнозов CATE	218
Эффект по квантилям модели.....	219
Кумулятивный эффект.....	224
Кривая кумулятивного прироста.....	225
Преобразование цели.....	228
Когда прогнозные модели хорошо упорядочивают эффекты.....	230
Убывающий предел отдачи.....	230
Бинарные результаты	231

Использование CATE для принятия решений	233
Основные выводы	236
Глава 7. Метаобучаемые системы.....	238
Метаобучаемые системы для дискретного воздействия.....	239
Т-обучаемая система.....	240
Х-обучаемая система	243
Метаобучение при непрерывном воздействии.....	248
S-обучаемая система.....	249
Двойное машинное обучение / машинное обучение с устранением смещения	253
Основные выводы	261

ЧАСТЬ IV

ПАНЕЛЬНЫЕ ДАННЫЕ

Глава 8. Разность разностей.....	264
Панельные данные	265
Каноническая разность разностей	268
Разность разностей с ростом результата	270
Разность разностей с методом наименьших квадратов.....	271
Разность разностей с фиксированными эффектами.....	273
Множественные периоды времени	273
Анализ.....	275
Предположения идентификации.....	278
Параллельные тренды	278
Предположение об отсутствии предварительных ожиданий и SUTVA	280
Строгая экзогенность	281
Отсутствие спутывающих переменных, изменяющихся во времени	282

Отсутствие обратной связи	283
Отсутствие переноса и запаздывающей зависимой переменной	284
Динамика эффектов во времени	284
Разность разностей с независимыми переменными	287
DID с двойной робастностью	290
Модель меры склонности	290
Модель дельты результата	291
Теперь все вместе	291
Поэтапное внедрение	293
Гетерогенный эффект во времени	299
Независимые переменные (ковариаты)	302
Основные выводы	303
Глава 9. Синтетический контроль	305
Набор данных онлайн-маркетинга	306
Матричное представление	308
Синтетический контроль как горизонтальная регрессия	310
Канонический синтетический контроль	315
Синтетический контроль с независимыми переменными	318
Устранение смещения в синтетическом контроле	322
Анализ	326
Синтетическая разность разностей	328
Снова о DID	328
Снова о синтетическом контроле	329
Оценка весов во времени	331
Синтетический контроль и DID	333
Основные выводы	335

ЧАСТЬ V

Альтернативные варианты дизайна эксперимента

Глава 10. Гео- и switchback-эксперименты	338
Геоэксперименты.....	339
Проектирование синтетического контроля	340
Проверка случайного набора объектов с воздействием	343
Случайный поиск.....	345
Switchback-эксперимент	348
Потенциальные результаты последовательностей	350
Оценивание порядка эффекта переноса	351
Оценка на основе дизайна	353
Оптимальный дизайн switchback-экспериментов	358
Робастная дисперсия.....	360
Основные выводы.....	363
Глава 11. Некомплаентность и инструменты	365
Некомплаентность.....	365
Расширение потенциальных результатов.....	367
Предположения идентификации инструментов	370
Первая стадия.....	373
Редуцированная форма.....	373
Двухшаговый метод наименьших квадратов	374
Стандартная ошибка	375
Дополнительные контрольные переменные и инструменты	378
Самостоятельная реализация 2SLS.....	379
Матричная реализация	380
Метод разрывной регрессии.....	382
Предположения разрывного метода	384
Эффект намерения к воздействию	384

Оценка ИП	386
Незапланированная группировка.....	387
Основные выводы.....	387
Глава 12. Что дальше	390
Выявление причинности	391
Последовательное принятие решений.....	391
Обучение с причинным подкреплением	392
Причинное прогнозирование.....	393
Доменная адаптация.....	393
Несколько слов напоследок	394
Об авторе	395
Иллюстрация на обложке	396